

1

AD-A052391

DEFENSE SYSTEMS

MANAGEMENT COLLEGE



PROGRAM MANAGEMENT COURSE INDIVIDUAL STUDY PROGRAM

A NEW, INTEGRATED APPROACH TO ARMY
OPERATIONAL TESTING AND EVALUATION

STUDY PROJECT REPORT
PMC 77-2

Freeman "Gordon" Lee
GS-13 OTEA

D D
RRCBII
APR 4 1978
RRCBII
D

FORT BELVOIR, VIRGINIA 22060

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NEW, INTEGRATED APPROACH TO ARMY OPERATIONAL TESTING AND EVALUATION		5. TYPE OF REPORT & PERIOD COVERED Study Project Report 77-2
6. PERFORMING ORG. REPORT NUMBER		
7. AUTHOR(s) FREEMAN "GORDON" LEE		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEFENSE SYSTEMS MANAGEMENT COLLEGE FT. BELVOIR, VA 22060		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DEFENSE SYSTEMS MANAGEMENT COLLEGE FT. BELVOIR, VA 22060		12. REPORT DATE 1977-2
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 81
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) UNLIMITED		
<div style="border: 1px solid black; padding: 5px; text-align: center;"> DISTRIBUTION Approved for _____ Distribution Unlimited </div>		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) SEE ATTACHED SHEET		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) SEE ATTACHED SHEET		

DEFENSE SYSTEMS MANAGEMENT COLLEGE

STUDY TITLE:

A NEW, INTEGRATED APPROACH TO ARMY OPERATIONAL TESTING AND EVALUATION

STUDY PROJECT GOALS:

The goals of this report are to provide the US Army with a logical basis to develop a data base, model, and technology for operational test and evaluation. The concept could be incorporated into AR 71-3 to provide a consistent policy for those organizations involved in operational testing.

STUDY REPORT ABSTRACT:

An approach is developed to evaluate operational effectiveness and military utility for operational test and evaluation. The technique is applicable to many areas where the systems' engineering approach involves both quantitative and subjective elements. Operational test organizational interactions are given in Section IV. A model is developed in Section V. Appendix B contains a technique to determine the operational effectiveness and military utility. Appendix C outlines the Delphi technique and Appendix D the application of utility theory to the problem of quantifying subjective information.

SUBJECT DESCRIPTORS

- Utility Analysis.....(10.01.02.02)
- Planning and Control Systems.....(10.02.05)
- Decision/Statistical Analysis Methods.....(10.04)
- Utility Theory.....(10.04.07)
- Judgmental Probability.....(10.04.09)
- Systems Engineering Management.....(10.05)
- Specialty Requirement.....(10.05.06)
- Test and Evaluation.....(10.08)
- Development Test.....(10.08.02)
- Operational Test.....(10.08.03)

NAME, RANK, SERVICE Freeman Gordon Lee, GS-13, US Army	CLASS PMC 77-2	DATE 4 November 1977
---	-------------------	-------------------------

A NEW, INTEGRATED APPROACH TO ARMY
OPERATIONAL TESTING AND EVALUATION

Individual Study Program

Study Project Report

Prepared as a Formal Report

Defense Systems Management College

Program Management Course

Class 77-2

by

Freeman "Gordon" Lee
GS-13 OTEA

November 1977

Study Project Advisor
Mr. Thomas F. Keegan

This study project report represents the views, conclusions and recommendations of the author and does not necessarily reflect the official opinion of the Defense Systems Management College or the Department of Defense.

EXECUTIVE SUMMARY

Decision Process

Implementation of the recommendations of this study project will have an impact upon the Army system acquisition decision process. Formal operational effectiveness/military utility prediction has not been accomplished in a systematic manner in the Army. Decision processes will tend to become more formalized and prescribed. The use of formal decision algorithms will become more widespread. This does not mean that the management decision process has been relegated to a witless computer. It does mean that management will have a new wealth of correlated facts quickly available, and the decision process will become easier and more accurate in many instances. Nevertheless, the ultimate act of decision must rest with a human who can account for the qualitative aspects of the world, those psychological and political intangibles that the formalized trade-off studies do not encompass.

Implement Decision

It should be carefully noted that the steps of the task analysis under discussion apply in any and all phases of a system life-cycle. Accordingly, implementation of a decision will tend to differ from phase to phase. In the conceptual phase the decision may take any of the following forms:

- (1) Further studies
- (2) Initiation of research, analysis and/or subtest
- (3) Initiation of exploratory development
- (4) Revision of an existing requirement or issue
- (5) Initiation of a new requirement or issue

In the validation and later phases, decision implementation tends to become more constrained.

Change Analysis

The implementation of a decision based upon operational effectiveness/military utility considerations generally implies a change in one or more of the following areas:

- (1) Schedule
- (2) Model(s)
- (3) System
- (4) Requirements

Each iteration of the operational effectiveness/military utility prediction/evaluation/augmentation cycle should be accompanied by a change analysis against these areas. The result of this activity will be a monitoring of the net effect of each decision and the accomplishment of program surveillance.

Principal Factors of Effectiveness

This study project takes the position that operational effectiveness is a quantitative measure of the extent to which a system may be expected to achieve a set of specific mission requirements (see Appendix B). It is expressed as a function of three major system attributes:

(1) Suitability is a measure of the operational condition of the total system (not just the hardware) at the start of a mission, when the mission is called for at an unknown (random) point in time.

(2) Dependability is a measure of the system condition during the performance of the mission; given its condition (suitability) at the start of the mission.

(3) Capability is a measure of the results of the mission; given the condition of the system during the mission (dependability).

It should be noted that this approach has a concept and definition of effectiveness based on quantifiable and subjective factors. There are certain aspects of the problem of effectiveness, and an effective military posture, which are purely psychological. An effective military posture is one which deters the enemy; or given that this does not occur, will abbreviate the conflict in favor of our national interest.

A well publicized threat of missile retaliation, backed in actuality by only a cleverly concealed squadron of "wooden missiles," might deter the enemy and satisfy the first half of the above requirement; but "wooden missiles" would not satisfy the second half of the requirement. However, it is difficult to quantify or assess the worth or value of deterrence. It must be left to military judgment. Appendices C and D of this study develop a technique to quantify these subjective factors. The difference between quantifiable factors (which are called risks) and non-quantifiable factors (called uncertainties) are discussed. Risk is akin to rolling dice or playing roulette. The outcomes are, on the average, quantifiable and predictable. Uncertainty is synonymous with lack of information or

inability to predict the outcome of the future; for example, the inability to prognosticate future weapon system configuration changes, either due to changes in hardware, operational concepts, or force size, and their consequent effect on costs. Uncertainty is a major factor in cost overruns.

Model

Once the relationships are established, an analytical model of the system can then be constructed. In the context of this study project, a model is any device, technique or process by means of which the specific relationships of a set of quantifiable system characteristics may be investigated. The advantage of using a model is that a wide variety of information may be employed to isolate problems within gross areas. Having done that it is possible to estimate the sensitivity of outcomes to variation of the parameters. This will permit the operational test designer to focus attention in areas of highest risk. This does not imply that one should design a test solely for evaluation in these areas, but it does suggest concentrating effort to validate findings. Adoption of the model approach would also entail the early establishment of a system data base from which all "team members" (i.e., developer, tester and user) will draw. Lastly, the association with the model would lead to earlier and better design of test strategy and improved assurance of the adequacy of data collection and reduction techniques.

Team Interactions

This study project outlines the interactions among the three main operational test and evaluation players - the Program Manager, the Test Manager and the TRADOC System Manager - and suggests ways to promote a symbiotic relationship between them. It also points out that the disciplines within the various functional domains are at different levels of development.

LIST OF FIGURES

Figure No.	Page No.
1. The Operational Test Iterative Process Over the Life Cycle	11
2. The Concept of Test and Evaluation in the Materiel Acquisition Process	14
3. An Integrated Systems View of the Operational Test and Evaluation Process	16
4. The Organization of Test and Evaluation During the Development Cycle	26
5. The Test and Evaluation Organizational System . . .	28
1B. Analytical Approach to Operational Effectiveness .	B-6
2B. Synthesis of Military Utility	B-7
1D. Minimum Acceptable Ratio for Military Utility . . .	D-6

LIST OF TABLES

Table No.	Page No.
1. Typical Footstones Associated with an Operational Test	36

ACKNOWLEDGEMENTS

The author is indebted to the many individuals that have contributed to, supported and encouraged me in this endeavor. I want to particularly express my appreciation to Mr. Tom Keegan, my Study Project Advisor, and to Mr. Wayne Schmidt, Professor of Industrial Management, DSMC, for their review, comments and patient consideration of my task. Additionally, I am indebted to MG Fred Ascani (USAF, Ret.), Professor, University of Southern California, COL Richard Rooth (USA), Chief, Science and Technology Division, USAOTEA, and Mr. Fred McCoy, Chief, Technical Operations Branch, Science and Technology Division, USAOTEA, for their review and helpful comments.

The outstanding artwork was the courtesy of the Graphics Section of USAOTEA.

The preparation of the final manuscript was performed by:

Ms Marge Grzesiak

Ms Barbara Howell

Ms Phyllis Collins

Ms Nancy Borg

The author lifted liberally from the excellent dissertation on models given in the Weapon System Effectiveness Industry Advisory Committee, Chairman's Final Report (Reference 2).

TABLE OF CONTENTS

EXECUTIVE SUMMARY	ii
LIST OF FIGURES	vii
LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
<u>Section</u>	
I. INTRODUCTION	1
Purpose of the Study Project	1
Approach to the Problem	3
Definitions	6
Scope of the Study Project	9
II. OPERATIONAL TEST AND EVALUATION	10
III. OT ORGANIZATIONAL INTERACTIONS	15
General	15
Organization, Maintenance and Personnel	20
Duplication	23
IV. TEST CYCLES	25
V. METHODOLOGY	31
General	31
Mission Definition	35
Resources	35
System Description	37
Figures of Merit	37
Specification of Accountable Factors	38
Identify Data Sources	39
Model Construction	39
Data Acquisition	40
Data Processing	41
Specify Schedule	42
Model Exercise	42
Prepare Management Summary Reports	46
VI. CONCLUSIONS	47

TABLE OF CONTENTS (Con't)

VII. RECOMMENDATIONS	49
APPENDIX A: LIST OF ABBREVIATIONS	A-1
APPENDIX B: PROPOSED TECHNIQUE FOR DETERMINING OPERATIONAL EFFECTIVENESS AND MILITARY UTILITY MILITARY . .	B-1
APPENDIX C: DELPHI TECHNIQUE	C-1
APPENDIX D: UTILITY THEORY	D-1
LIST OF REFERENCES	

SECTION I

INTRODUCTION

Purpose of the Study Project

The legislative basis for operational testing and evaluation as set forth by Congress is given in Section 139, Chapter 4 of Title 10, United States Code (as amended by the FY 1974 Authorization Act):

"The Secretary of Defense shall submit to Congress each calendar year. . . a written report . . . for each weapon system . . . for which any funds for procurement are requested in that budget. The report shall include data on operational testing and evaluation for each weapon system."

It is evident from the above quotation that Congress desires a basis for judging the adequacy of our defense posture and rationale for making management decisions. Operational testing and evaluation reports form an important part of this decision making process.

Because of the importance placed by Congress on operational testing and evaluation and the high costs associated with conducting field tests, increased emphasis must be placed in this area. However, the accelerated pace of design, development and obsolescence of military systems in recent years has given rise to a series of problems in systems management, many of which are addressed in Reference 1. For example:

1. Many of today's systems have a high unit cost. In the interest of economy, if the national budget is to be held in line with national objectives, there must be a clear indication of both the cost and effectiveness of a proposed system long before the decision is made to produce the system and put it into operational use. Thus, there is a clear need to

predict system effectiveness and life cycle costs as early in the system acquisition process as possible.

High unit costs also lead to abbreviated test programs during the acquisition phase. Consequently, there is a large degree of uncertainty in the quality of the product to be placed in the operational inventory. Better methods of quantification are needed to reduce this uncertainty.

2. Many of today's weapon systems tend to be "one shot" devices. As a result, the operational test must be designed around high risks or pay the price of scoping a test to an acceptable confidence level. There will be less direct, advance evidence as to the adequacy of the system. It is becoming increasingly necessary to rely on indirect evidence to focus attention on critical areas for verification and assurance of effectiveness.

3. Very few deny the necessity of defense. Yet, in the past few years there has been ever greater emphasis to reduce peacetime defense costs. At the same time there has been additional emphasis to increase wartime effectiveness. Maximizing effectiveness and minimizing costs at the same time is not logically possible. Thus, the real problem is to obtain as efficient a defense posture as possible within the constraints of cost and effectiveness, or stated another way, to optimize cost when the effectiveness is constrained or to optimize effectiveness when the cost is constrained. This is generally referred to as cost-effectiveness optimization. Optimization seeks to allocate the national resources in a way that withstands the critical vision of hindsight.

This is an extremely difficult problem since the defense posture is developed in the presence of risk and uncertainty. Thus, it is essential to seek and use the best available methods for cost effectiveness optimization.

4. Many Army programs and studies have been conducted under the name of system effectiveness, operational effectiveness, military utility, operational readiness, and like terms, but which all referred to the same problem. A need exists to focus attention in this problem area in order to standardize definitions, terminology, and a method of attack.

5. Finally, there is the problem of establishing quantitative requirements for complex systems, particularly when those requirements must be stated in very general or in probabilistic terms.

The purpose of this study project is to outline the current status of operational test and evaluation in the system acquisition process, and to provide a method for improving the confidence level associated with independent operational test evaluations. This is accomplished by identifying those areas that require additional development and directing a coordinated approach with an improved technique to reduce the risk currently associated with management decisions.

Approach to the Problem

Given adequate resources to conduct an operational test and subsequent evaluation in a preferred manner, the risk would be reduced to a point of being negligible. However, because of factors such as the high cost and the length of time involved in operational test, the difficulty

in acquiring a suitable environment and the practical problems with setting up the test, a risk higher than desired is almost inevitable. This leads one to ask "are there alternatives to testing?" The answer is "yes", but not very good ones. So, since tests are required, then how might one enhance the test results by other means - THIS IS WHAT THIS STUDY PROJECT IS ALL ABOUT!

A potential solution to the above stated difficulties is the judicious use of analytic modeling techniques to aid both in establishing subsystem requirements before development commences, and to compute the odds for mission success from less than full system test data.

In effect then, one is forced into the position of performing an analytic modeling program by default. Adequate system test and evaluation cannot be accomplished in any other practical way. The approach to modeling here is similar to that taken in Reference 2.

From the preceding considerations, the general role of analytic modeling is clear. Analytic models provide insight. They make an empirical approach to system design economically feasible. They are a practical method of circumventing a variety of exterior constraints.

One has a right, then, to expect certain kinds of output from a modeling program. Clearly, a modeling program should:

1. Aid in establishing requirements.
2. Provide an assessment of the odds for successful mission completion.
3. Isolate problems to gross areas.

4. Rank problems in their relative seriousness of impact on the mission.

5. Provide a rational basis for evaluating and selecting between proposed system configurations and proposed solutions of discovered problems.

Clearly, these outputs can be realized only if the scope of the modeling effort is adequate and only then when it is supported by a reasonable data base. Furthermore, these outputs are achievable only when the words "operational effectiveness" convey a definite meaning of sufficient scope.

The concept of operational effectiveness has been expressed many times in many ways by many people. Sometimes one characteristic, such as reliability, has been emphasized as a major contributor to operational effectiveness. At other times, other characteristics have been singled out for special attention.

The time has come to concentrate attention on the primary concern of management -- the overall operational effectiveness of a system -- and to derive a way to predict and measure this overall effectiveness and to put each contributing characteristic in its proper perspective within the overall measure.

A consistent method for modeling both operational effectiveness and military utility is given in Appendix B. A technique for quantifying the subjective elements is developed in Appendices C and D.

This study project also outlines the interactions among the three main players -- the Program Manager, the Test Manager, and the TRADOC

System Manager -- and suggests ways to promote a symbiotic relationship among them.

Definitions

Criticality to mission evaluation. One of the major tools for minimizing the test program without compromising mission integrity is the utilization of criticality evaluations of the hardware. Essentially, this technique evaluates the mission effects of hardware failure modes and establishes criticality indices as a function of probable mission success. Multiple failure events are also considered to provide as complete a failure derivation as possible. The amount of testing required to verify the integrity of a given hardware item is geared to the criticality category (required confidence level).

Decision Coordinating Paper (DCP). The principal document to record essential system program information for use in support of the Secretary of Defense decision-making process at Milestones I, II and III.

(Reference DOD Directive 5000.2)

Defense System Acquisition Review Council (DSARC). An advisory body to the Secretary of Defense on major system acquisitions. The Council members are the OSD staff principals. (Reference DOD Directive 5000.2)

Life cycle cost. Is the sum total of the direct, indirect, recurring, nonrecurring, and other related costs incurred, or estimated to be incurred, in the design, development, production, operation, maintenance and support of a major system over its anticipated useful life span.

(Reference OMB Circular A-109)

Major system. Is that combination of elements that will function together to produce the capabilities required to fulfill a mission need. The elements may include, for example, hardware, equipment, software, construction, or other improvements or real property. Major system acquisition programs are those programs that (1) are directed at and critical to fulfilling an agency mission, (2) entail the allocation of relatively large resources, and (3) warrant special management attention. Additional criteria and relative dollar thresholds for the determination of agency programs to be considered major systems under the purview of this Circular, may be established at the discretion of the agency head.

(Reference OMB Circular A-109)

Mission Element Need Statement (MENS). A statement prepared by a DOD component to identify and support the need for a new or improved mission capability. The mission need may be the result of a projected deficiency or obsolescence in existing systems, a technological opportunity, or an opportunity to reduce operating cost. The MENS is submitted to the Secretary of Defense for a Milestone 0 decision. (Reference DOD Directive 5000.2)

Operational Test and Evaluation (OT&E). Test and evaluation conducted to estimate the system's military utility, operational effectiveness and operational suitability. (Reference DOD Directive 5000.3)

Program objectives. Are the capability, cost and schedule goals being sought by the system acquisition program in response to a mission need. (Reference OMB Circular A-109)

(Service) System Acquisition Review Council ((S)SARC). A Council established by the Head of a Military Department as an advisory body to him and through him to the Secretary of Defense on major system acquisitions. The (S)SARC is chaired by the Secretary/Under Secretary of the Military Department and is similar in functional composition, responsibilities and operation to the DSARC. In application the term (Service) is replaced by the designation of the applicable Military Department, i.e., ASARC, NSARC and AFSARC. (Reference DOD Directive 5000.2)

System. Is an arbitrary collection of physical configurations together with the functions performed upon them. It is completely defined, when and only when, a set of influencing configurations and functions, called the environment, are given.

System Acquisition Process. A sequence of specified decision events and phases of activity directed to achievement of established program objectives in the acquisition of Defense systems and extending from approval of a mission need through successful deployment of the Defense system or termination of the program. (Reference DOD Directive 5000.1)

System Deployment. Delivery of the completed production system to the using activity.

Survivability. Is the probability that a system will either (1) be removed from the threatened environment before it can be attacked (as with warning), or (2) "ride out" some anticipated attack.

Vulnerability. Is that characteristic of a system which causes it to suffer a definite degradation (incapability to perform the designated

mission) as a result of having been subject to a certain level of effects in an unnatural (man made) hostile environment.

Scope of the Study Project

This study project is designed to aid the developer, tester, and user in accomplishing operational test and evaluation that is efficient as well as adequate. This synergistic process must be initiated and a common data base started at an early date in the system life cycle in order to be most valuable to the players. There was no attempt to make this an exhaustive study; only to identify the problems, suggest a solution, and provide the Program Manager, Operational Test Manager and the TRADOC System Manager with a means to minimize the risks and uncertainties associated with program decisions. References are provided for those individuals seeking more detail.

SECTION II

OPERATIONAL TEST AND EVALUATION

Operational Test (OT) is testing conducted by military personnel to determine the degree to which new systems fulfill military operational requirements. It is conducted under conditions which duplicate as closely as practicable the environment expected in field operations. The system addressed by OT includes not only the hardware but also the personnel and means of employment. Therefore, OT includes the maintenance and logistic support, as well as the operation of the equipment, and the examination of training, tactics and techniques for most effectively using the system in combat.

The primary purpose of operational test and evaluation is to provide information for decision making. Specifically, operational test and evaluation provides information on how well a development program is meeting the system objectives and what the ultimate outcome of the program is likely to be. Since this is primarily concerned with predicting how well the system will perform once it becomes operational, it is essential to most major decisions made during the course of a program. OT is essentially an iterative process or series of phases over the life cycle as portrayed schematically in Figure 1. Each iteration is keyed to an appropriate decision point. A test data base is established and updated as additional information becomes available.

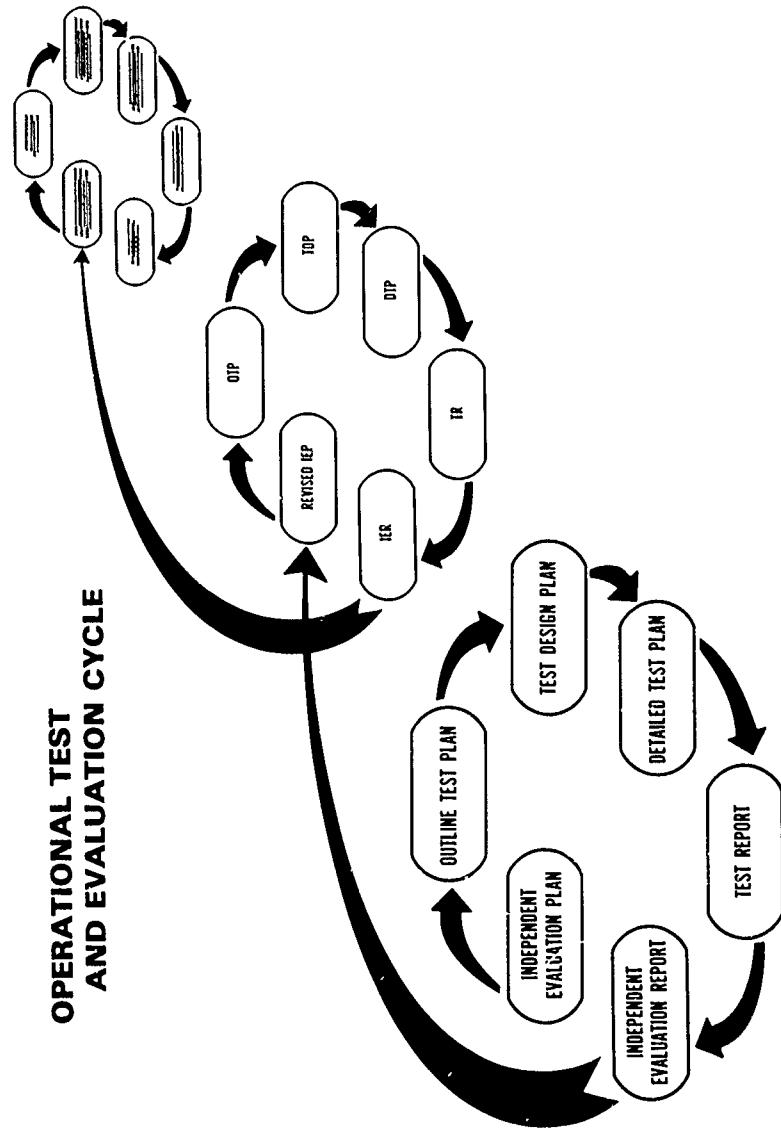


Figure 1. The Operational Test Iterative Process Over the Life Cycle.

The confidence that can be attached to the information provided by operational test and evaluation is directly related to the amount of testing accomplished (assuming a well designed test) i.e., the more testing accomplished, the more confidence there should be in assessing the progress and likely outcome of the program. However, operational testing is expensive and a judgment must be made as to how much confidence is required, or conversely, how much risk can be accepted in program decisions.

OT has a direct relationship to the military capabilities of our operating forces in the field. The more effort expended to identify and correct operational deficiencies before the system is placed in production, the greater will be the military capability of the operating forces when the new weapon system is deployed. Significant cost benefits may be realized when major system deficiencies are corrected prior to operational deployment. Modification programs (retrofitting) applied to systems after they have entered the operational inventory are frequently extremely expensive, time consuming, and cause non-availability of the system.

The Project Manager, being acutely aware of his responsibility to field a system acceptable to the user, must take advantage of every opportunity to make the most effective use of the test effort associated with his program. This not only means to develop a Test and Evaluation Master Plan (TEMP) that minimizes duplication in test, but to be constantly alert to situations that may lead to more effective use of the test resources available to him. The Army currently meets the requirements for the TEMP with the test and evaluation portion of the Outline

Development Plan (ODP) and the Coordinated Test Program (CTP). Reference to the TEMP in this report incorporates this concept.

The operational test and evaluation process starts with approval of the Mission Element Need Statement (MENS) by the Secretary of Defense, establishing Milestone 0 in the materiel acquisition cycle. This process is shown schematically in Figure 2 and discussed in detail in Section IV.

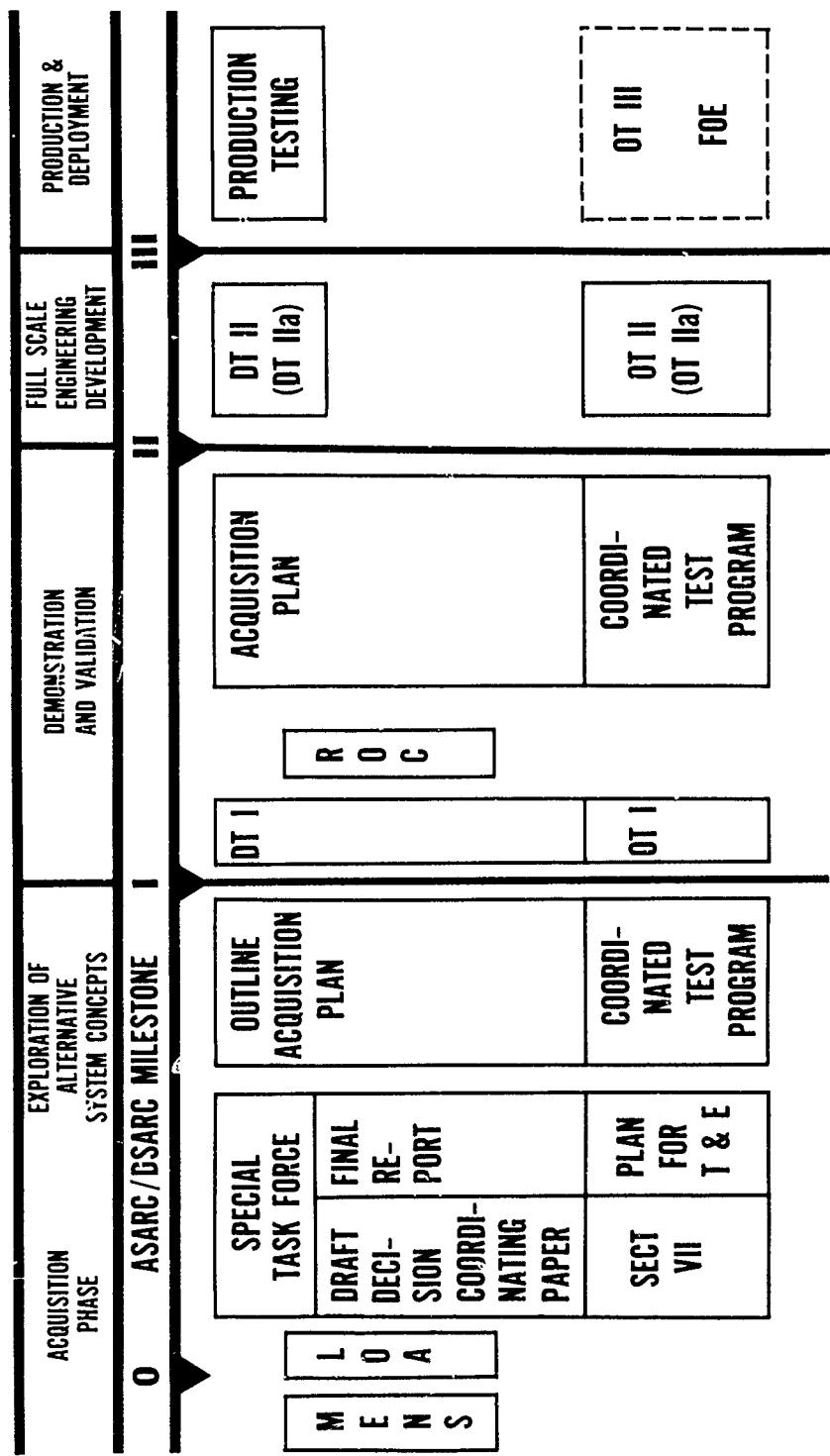


Figure 2. The Concept of Test and Evaluation in the Materiel Acquisition Process.

SECTION III

OT ORGANIZATIONAL INTERACTIONS

General

The principle organizations associated with the US Army Operational Test and Evaluation function are: the Project Manager's Officer (PMO); the user organization - Training and Doctrine Command (TRADOC); and the Operational Test and Evaluation Agency (OTEA). When the Secretary of Defense approves the program initiation at Milestone 0, the Department of the Army will assign a Program Manager (PM) for a major system acquisition. The PM will be given a charter, approved by the Secretary of the Army, starting the PM's responsibility, authority and accountability for program objectives. The Commanding General, TRADOC, appoints a TRADOC Systems Manager (TSM) who is the single point of contact for the user on this system. The Commanding General, OTEA, appoints a Test Manager (TM) also during this time frame, to manage the operational test and evaluation during the system acquisition process. For nonmajor systems tested and evaluated by the user, the Army Communications Command, and/or the Surgeon General, the participants and processes are similar but at lower echelons and called by different terms.

One approach to viewing the integrated operational test and evaluation process is shown in Figure 3. This shows the system in the operational environment suprasystem, interacting with the hardware subsystem, the maintenance and logistics subsystem, personnel and training subsystem, organization and doctrine subsystem, and the operational evaluator. The system is prescribed in DOD Directive 5000.1, Major System Acquisitions

THE OPERATIONAL EVALUATION SYSTEM

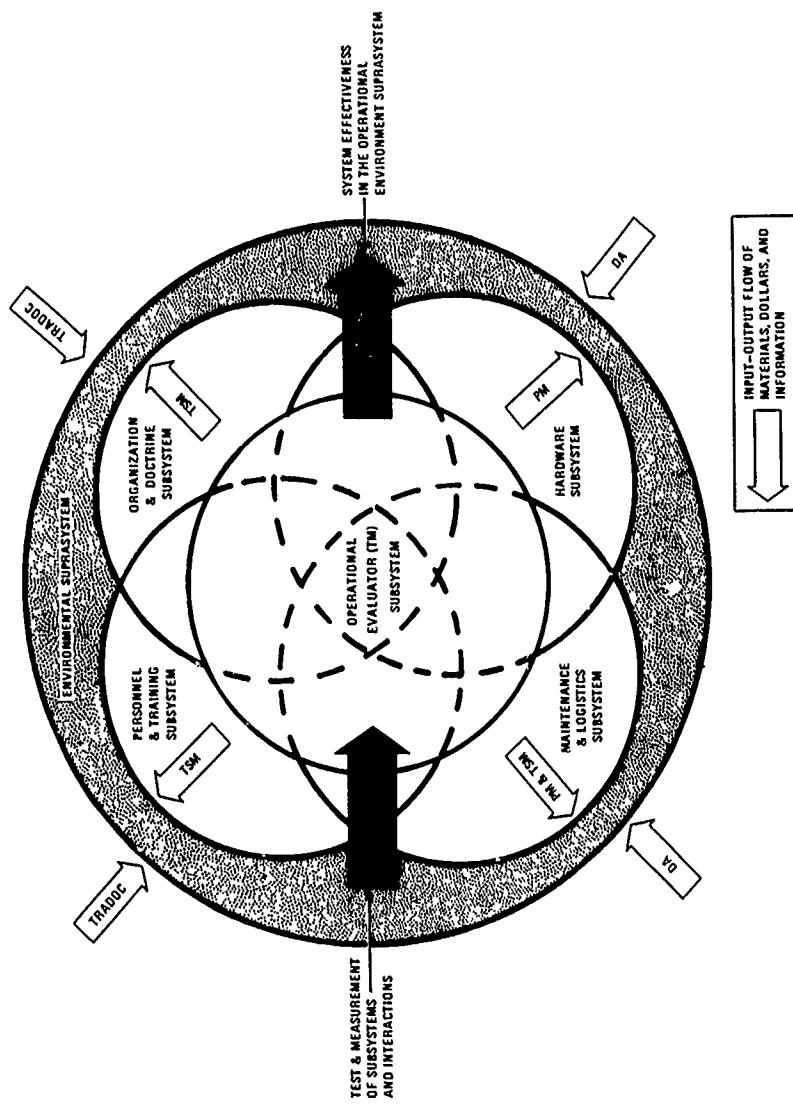


Figure 3. An Integrated Systems View of the Operational Test and Evaluation Process.

(18 Jan 77), which states:

"Test and evaluation shall commence as early as possible. An estimate of military utility and operational effectiveness and operational suitability including logistic support requirements, shall be made prior to large scale production commitments. The most realistic test environment possible and an acceptable representation of the future operational system will be used in the testing. . ."

For purposes of discussion the diagram (Figure 3) shows graphically, that in a typical operational test and subsequent evaluation of a whole system, that four major factors drive the system operational effectiveness: hardware effectiveness, maintenance and logistics effectiveness, personnel and training effectiveness, and organization and doctrine effectiveness. These factors are not necessarily equal in weights in determining the operational effectiveness and the current status of methodology associated with each varies greatly. For example, a scientific method of evaluating the parameters of effectiveness in maintenance and logistics, personnel and training, and organization and doctrine, has neither been developed nor the major characteristics defined. It is estimated by the author that for a complicated system the degree of development of the operational test and evaluation methodology for the effectiveness of each of the four major subsystems is:

Hardware	90-95%
Maintenance & Logistics	10-20%
Personnel & Training	5-10%
Organization & Doctrine	45-50%

The indication here is that the techniques for determining and evaluating the hardware characteristics are almost complete; however, in the other cases neither the major (driving) characteristics have been fully identified nor techniques for testing and evaluating them have been adequately developed. Consequently, these items are addressed in the most general terms (e.g., test and evaluate the logistics system) and tested by exception (e.g., list all logistical shortcomings observed during the conduct of the test). This limits the confidence level expected from testing in each of these areas (e.g., in order to obtain a 0.8 "reliability" for this system, each of the four subsystems would have to obtain a 0.947 reliability).

Reference 3 cites a typical example. In June 1974, an engagement test of US Army armor was conducted in Wildflecken, Germany, pitting a tank company against an infantry force equipped with TOWs. A simple test of target acquisition and gunnery was used, employing telescopes and large numbers on the participant TOWs and tanks. The test required the TOW force to engage the tanks in repetitive contests over the same ground. In the final "battle", losses sustained by the tankers were 33 percent less than in the first. Assuming the validity of the test, the tankers had been taught how to use the terrain to better advantage in coping with a long-range guided missile system. Roughly, this improvement could be equated to providing each US Army tank company so trained with an additional tank platoon during its final engagement with the enemy. Or, the \$1600 invested for training devices to train the Wildflecken company could be said to have returned \$1,600,000, the

procurement cost of a tank platoon; i.e., 1000:1. It should be noted that the driving characteristics leading to this discovery were not identified and further evaluated to determine the relationship. Result - more trial and error.

The Wildflecken test was also interesting in that the participating tank company was considered "well trained" when it started. Its heavy losses in the initial iterations of the engagement test brought consternation to the colonel commanding, which gave way to elation when he perceived the tactical proficiency being developed in each successive run. But that colonel had seriously underestimated the vulnerability of his tanks to modern weaponry and just such miscalculation occasioned exorbitantly high losses among Israeli armor early in the Yom Kippur War. The US Army cannot afford to fight the first battle of the next war with loss rates comparable to Israeli attrition. If so, it could lose within a few days, more tanks than we now have in Europe. But the current state of training methodology does not allow a direct approach to identifying the characteristics that really are the basic issues; the issues must be evaluated indirectly by conducting a mini war and hope that any shortcomings will be observed.

The above discussion is presented to emphasize the fact that the confidence one has in the test results in each of these areas should vary considerably, and any attempt to combine these into an overall system index should be carefully evaluated.

Organization, Maintenance and Personnel

The Army does not have an adequate data base in these areas (see Figure 3) against which to measure operational effectiveness. Such a data base for current weapon systems must be developed to provide a basis for evaluating new concepts and marginal increases in effectiveness promised by new systems. There is neither evidence that the requirement for such a data base has been developed nor are required characteristics and information being developed on a systematic basis. Further, more work should be directed toward the early assessment of innovative ideas dealing with tactics and doctrine.

The engineering community has quantified the Reliability, Availability and Maintainability (RAM) in the hardware subsystem; why not "RAM" in the other three subsystems? This would require a "hard look" at each of the subsystems to perceive what are the driving characteristics of each and devise means of quantifying those that are subjective. For example, we spend millions of dollars to build expensive simulators to train the troops in as close to identical systems as possible, without ever questioning what are the critical characteristics in the operation. In a recent foreign purchase a country declined to use our expensive training simulator, substituted a small box with a manual operation, and obtained better operational effectiveness than we had with the simulator.

The Army hadn't perceived the important concept of operation as related to the user. Other countries maintain higher availability with less maintenance, on identical equipment, by changing the doctrine and tactics.

Relationships must be developed between characteristics in order to apply the systems engineering approach. For example in maintenance, the system model ought to do more than identify the "best" alternative . . . it ought to contribute to the efficiency of the life cycle and logistic support process. It must be flexible enough to identify those areas where a large percentage of replacement actions are attributable to a small number of individual posts while acknowledging that the number of requisitions are always higher than replacements. It is interesting to note that while organizational personnel perform over 90 percent of the total replacement actions recorded, the direct man hours required for most specific tasks are considerably less than those required for direct support unit replacement operations. In general, it appears that organizational manpower does the same job with equal or less man hours compared to the direct support unit. In some cases engineering dictates decisions as to what must be repaired at site, direct support, general support, depot and/or the factory, and how it must be transported. In many other cases, however, repairable components can be handled at any location. The model must provide a basis to answer the major question of which location is best for which items - in terms of cost and effectiveness. Inherent in the repair, location decision is the type of transport system to be used.

In other cases, certain changes can be predicted without being able to predict the reasons for them. Such is the case in connection with the military significance of outer space on the future systems deployment and doctrine. It seems a reasonably safe conclusion, that by the time of the Initial Operational Capability (IOC) of systems currently in the validation

phase, military operations in outer space will be of profound importance to the survivability of the system. However, it cannot be determined if this significance will lie in a reconnaissance capability, offensive or defensive weapon capability, or in a command and control capability.

It would be extremely helpful to the operational evaluator if issues in all four subsystems were developed for the operational tester. The TSM (see Reference 4 for responsibilities) should attempt to devise characteristics to be validated in Force Development Testing and Experimentation (FDTE) and OT. This would be the start of a data base that would eventually lead to the "real" driving characteristics in these subsystems.

Actions. PM should obtain a dedicated human factors engineer at the start of the conceptual phase. This individual would be responsible for identifying areas where the maximum savings in the operations and support costs and manpower might be accomplished in the system life cycle. He would also be an excellent individual to interface with the TSM as they both have common goals. The human factors engineers would also be the repository for data relating to the man-machine interface within the PMO.

TSM should task the appropriate school or TRADOC component (e.g., TCATA, CDEC, AMSAA and/or TRASANA), at least six months prior to the requirement to submit issues to the operational tester and to determine the major characteristics in each of the three areas of TRADOC responsibility (see Figure 3). These characteristics should be ranked in their order of priority and sent with the issues to the operational tester. The schools should be encouraged to development system models for each of their

respected areas and maintain a data base.

TM should have the Evaluation Division (OTEA) design an Independent Evaluation Plan that addresses these characteristics. The plan should require information and operational test data that will lead to quantitative values for the operational effectiveness, operational suitability and military utility for each of the TRADOC subsystem areas in Figure 3. These values may not be absolute but must at least be comparable to the data base.

Duplication

Developers, users, and testers have expressed considerable difficulty in sorting out the proper division of testing responsibilities between OT and Development Test (DT). There is an actual difference both in concept and execution between DT and OT, but there remains undesirable duplication, resulting primarily from failure of the Test Integration Working Group to fully explore all areas where there is a commonality of data. There are two areas that appear to provide the best payoff: (1) Those Engineer Development Test (EDT), DT and OT where small changes would provide common data, e.g., by employing military personnel with appropriate MOS (instead of technicians) in human factors subtests, and (2) the increased utilization of computer and simulator results to aid in determining where to place emphasis in the OT test design. The independence between user and developer philosophy is pertinent and necessary; however, the emphasis should be changed from separate testing to independence of design and evaluation to permit more efficient use of testing resources applied to integrated or combined tests. The time to do this is during the DT

and OT test design processes; to attempt to aggregate after the fact is not usually possible.

Actions. PM should call a one day meeting with the TSM (courtesy and information), TM, contractor and PMO test personnel while planning the test program (prior to the formation of the TIWG). At this meeting each area of test and simulation should be reviewed to determine if there are areas that may or could produce data to support OT. During the Validation Phase the operational tester has very little data to support the OT I and in later phases he needs all the data he can accumulate to aid in the test design and evaluation. This is particularly true in the case of human factors and logistics. The operational tester may be in a position to complement some engineering and/or development testing and thereby reduce overall test (and costs) requirements.

TM should not miss any opportunity to obtain data of an operational nature. This may be in the form of early human factor engineering checks to detailed operational simulations. The information obtained for the system and certain assemblies can be of great use to the test designer and to support in many cases, the independent evaluator. The principle to promote with the PM is mutual trust and helpfulness.

TSM should use this opportunity to become better acquainted with the system and the "team".

SECTION IV

TEST CYCLES

The operational test and evaluation cycles in the materiel acquisition process are illustrated in Figure 4 and shown in detail in Figure 2. This is in accordance with DOD Directive 5000.3 (19 Jun 73, Chg 2, May 75) and the TEMP, and is required

". . . as early as possible in the acquisition process, and prior to initiation of Full Scale Development. . ."

OTEA has the responsibility for overall Army management of operational test. In performing this mission, the operational tester has the responsibility for operational test on all major and selected non-major systems, with the user representative performing those on most others. (Army Communications Command and the Surgeon General conduct about 5% of the non-major tests.) In addition, OTEA reviews and/or monitors all operational test designs prepared by TRADOC. Test design for major and selected non-major systems is accomplished by the operational tester; and by the TRADOC Test Boards for most other systems. The TRADOC Combined Arms Test Activity (TCATA) and the Combat Developments Experimentation Command (CDEC) can, as required, perform test design either for OTEA or for TRADOC. Conduct of the tests is accomplished either by OTEA or by the TRADOC Test Boards, however, OTEA because of its limited size, requires augmentation from TRADOC assets (TCATA, the Test Boards and, as required, CDEC) and the Forces Command (FORSCOM) troop support to conduct the tests. Test evaluation is performed by both OTEA and by the TRADOC schools and centers.

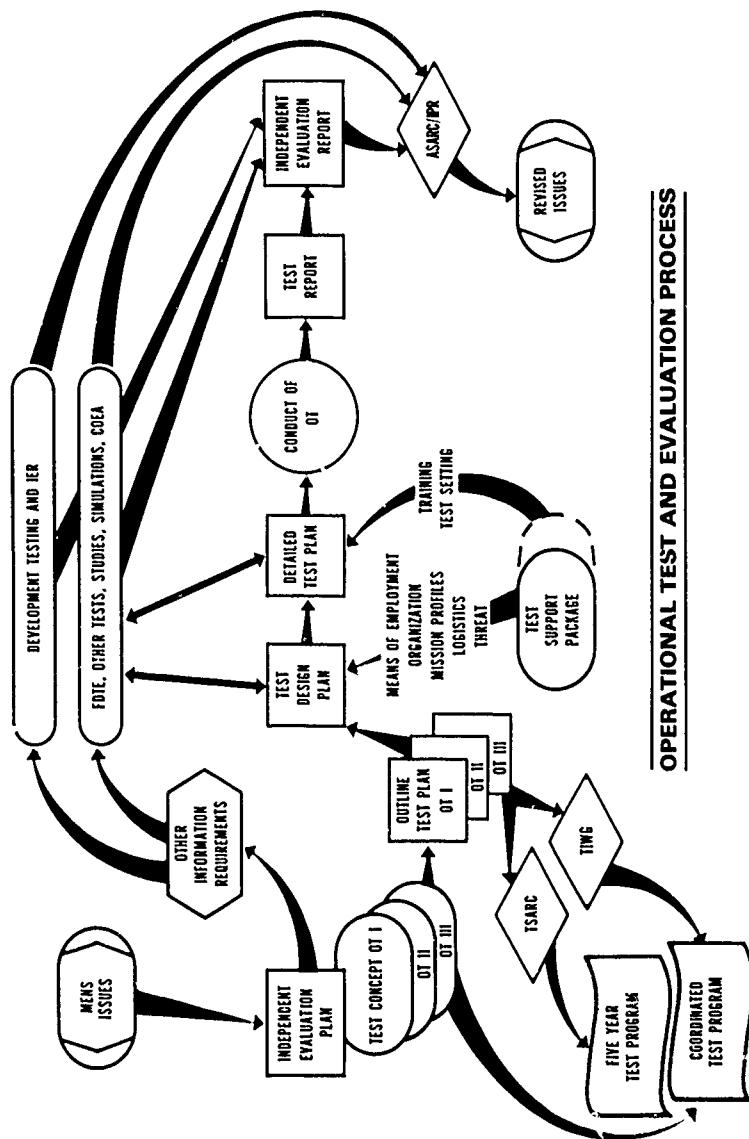


Figure 4. The Organization of Test and Evaluation During the Development Cycle.

This process is shown in Figure 5. The Logistics Evaluation Agency (LEA), which acts as the overall logistician during the materiel development process, evaluates the results of both DT and OT. DT/OT results are then presented to the Army and Defense Systems Acquisition Review Councils (ASARC/DSARC) for major systems or to an In-Process Review (IPR) for non-major systems. Attendees at the IPR include TRADOC, Materiel Development and Readiness Command (DARCOM), LEA and OTEA. During this process, the continued validity of the systems is checked. If still valid, the development process continues to the next more detailed phase until the development is complete and production and deployment initiated.

The above discussion has outlined the separate DT and OT processes. The next paragraph shows how the DT/OT system is integrated and how selected other organizations augment the system.

Test integration within DT means literally accepting contractor tests in lieu of Government testing. Whereas, integration of DT/OT means:

1. Maximum sharing of prototypes. Day-to-day test schedules of both DT and OT may be carefully inter-scheduled to take advantage of limited prototype availability.
2. Concurrent DT and OT. Although DT conceptually precedes OT, some overlap in timing can shorten the material development time.
3. Combined DT and OT where feasible. With proper coordination and planning, a subtest can produce data for both DT and OT simultaneously.
4. Aggregation of data but not the replacement of OT by DT. When appropriate, reliability data or mileage data may be aggregated to larger totals than either tested alone.

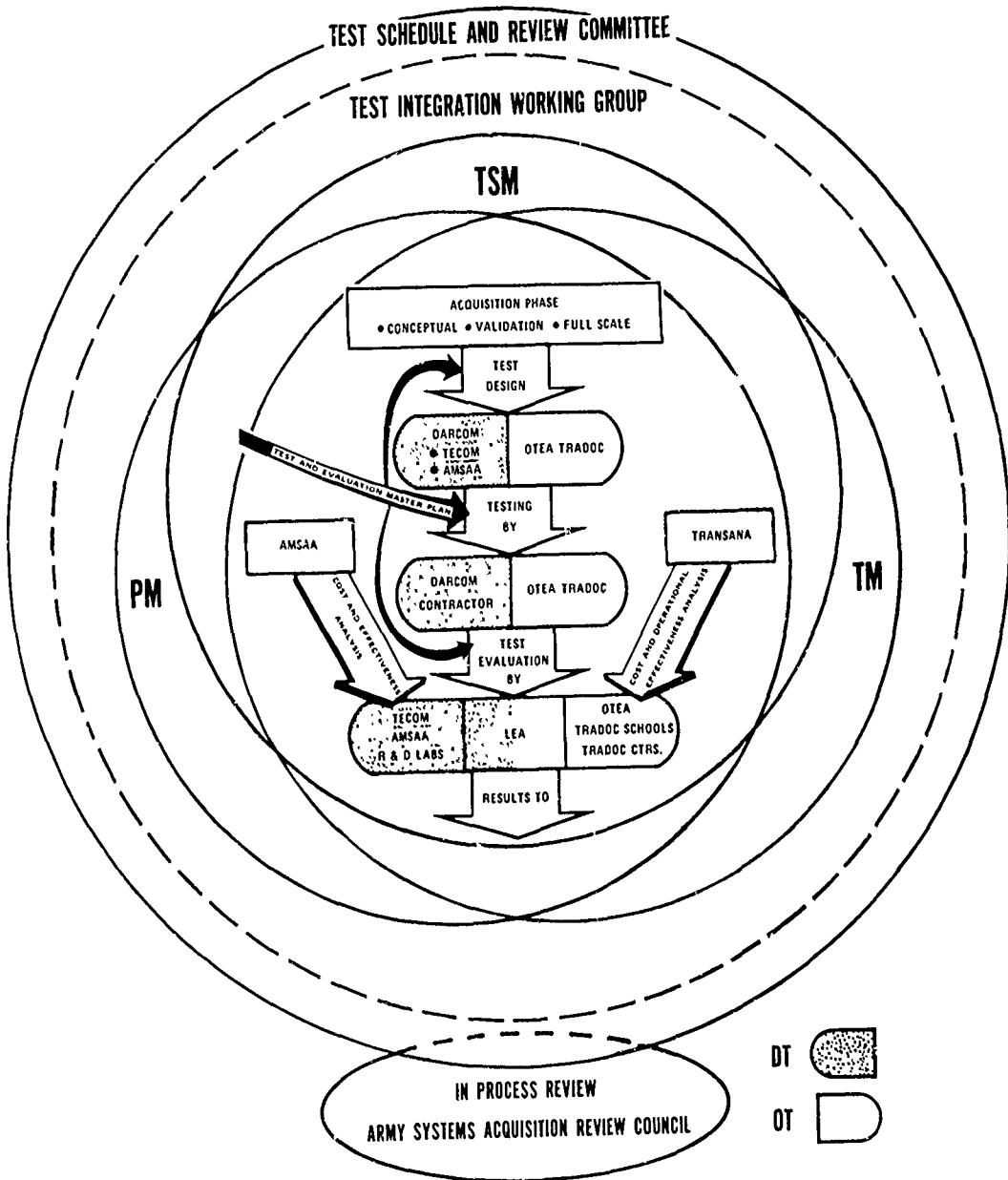


Figure 5. The Test and Evaluation Organizational System..

The following organizations have been set up to augment and to integrate the DT/OT process. Army Materiel Systems Analysis Activity (AMSAA) provides cost, effectiveness, tradeoff, and other inputs to the DARCOM developers and, as required, to TRADOC and OTEA. TRADOC has Army-wide responsibility for performing materiel/weapons systems Cost and Operational Effectiveness Analyses (COEA). Within TRADOC, the Training and Systems Analysis Agency (TRASANA) has a major role in accomplishing or supporting COEA development. These COEA's are provided to both OTEA and to the TRADOC centers and schools for use during the evaluation process.

The set of organizations shown in the outer rings of Figure 5 are designed to aid the Program Manager in that the DT/OT aspects of the development process are integrated.

1. The Test Scheduling And Review Committee (TSARC) is chaired by OTEA. The committee functions to produce the Five-Year Test Plan. The plan insures that weapon system testing is integrated and coordinated and minimizes the impact on the readiness of Forces Command units which are normally used for operational testing.
2. The DARCOM Project Manager's function is to insure that development problems which arise are solved and in conjunction with the newly authorized TRADOC Systems Manager to insure that user-developer coordination occurs on a daily basis.
3. The Test Integration Working Groups (TIWG) are composed of:
 - a. Principals. (those commands/activities with responsibilities and with whom responsible commands/activities must coordinate.)

- (1) Materiel Developer.
- (2) Combat Developer.
- (3) Operational Tester.
- (4) Development Tester.
- (5) Development Test Evaluator (if different from DT tester).
- (6) Operational Test Evaluator (if different from OT tester).
- (7) Contractor (where and when appropriate).
- (8) Logistician.
- (9) User (if different from combat developer).
- (10) Trainer (if different from combat developer).
- (11) Other commands/agencies/services (when appropriate).

b. Associates. (Those commands/activities who serve in a monitor's role; i.e., surgeon general representative for health aspects associated with testing, and/or subcommands/activities of principal attendees.)

Associates attend TIWG meetings in an advisory role.

TIWG's are established for selected systems. Their function is to convene on an "as required" basis to coordinate DT and OT testing, insure exchange of appropriate data, and to insure that duplication of test efforts does not occur.

SECTION V

METHODOLOGY

General

The Department of the Army (DA) has determined that materiel acquisition programs have routinely been conservatively designed with low risk strategies that required completion of each phase of the life cycle model and its associated testing. In conformance with the desire to develop the most effective acquisition program for each system, the following principles should be applied in the formulation of test programs:

1. DA has strongly reconfirmed the guidance of DOD Directive 5000.1 and AR 1000-1 that no single formula applies to all materiel acquisitions. Test and evaluation programs must be flexible and tailored to support the acquisition strategy for the system/item; (Reference 5 will aid in achieving this end) as opposed to routinely following the idealized life cycle test model depicted in AR 70-10 and DA PAM 11-25.
2. Life cycle test programs will be designed to match the acquisition strategy of a given system and altered as permitted or required by the results of testing. Scheduled test programs will be examined at Milestone I (Validation) and at Milestone II (Full Scale Engineering Development) to determine if expanding the test in the next phase would provide necessary information earlier, thereby reducing the overall period that must be devoted to testing throughout the total acquisition cycle. For example, procuring more prototypes could provide more test data earlier. Production facsimile materiel procured as test prototypes for DT II and OT II may

demonstrate in testing that the materiel systems meet their technical and operational requirements and confirm producibility when tested, thereby permitting reduced testing, following Milestone III (OT III and tests during Production/Deployment phase).

3. Testing of materiel during the Validation Phase will include DT I and OT I. Testing during the Full Scale Engineering Development Phase will include DT II and OT II. If the acquisition strategy calls for a production and deployment decision at Milestone III, DT II and OT II must provide the data for a valid estimate of the system's military utility, operational effectiveness and operational suitability (including compatibility, interoperability, reliability, availability, maintainability, logistic supportability, man (soldier) machine interface and training requirements). If the Milestone III decision is for production and subsequent testing include production testing (e.g., First Article Tests) by the developer and Follow-on Evaluation (FOE) by the operational tester or Initial Operational Capability-Force Development Testing and Experimentation (IOC-FDTE) by the user representative. If the Milestone III decision is for Low Rate Initial Production (LRIP), subsequent testing includes sufficient DT III integrated with production testing by the developer and OT III by the operational tester to support a full scale production decision. Initial production testing (e.g., First Article Preproduction Testing) following LRIP can serve to reduce production testing after the full scale production decision.

4. The ASARC/IPR decision at Milestone III addresses the system's readiness for transition into production and deployment. Once the

decision for full scale production in any quantity is made in order to provide materiel for the inventory, the materiel will be type classified as Standard. LRIP is a production decision option that should be exercised when the risks of a full scale production decision are unacceptable from both technical and operational considerations. Circumstances for LRIP include situations where producibility (i.e., progressing from the engineering prototype to the manufactured production item) is a significant risk. LRIP materiel procured for DT III and OT III will be type classified as Limited Production (LP). Production of items type classified LP for other purposes (e.g., emergency procurement) will be approved by HQDA on a case-by-case basis.

5. OT will be planned, coordinated with DT and reported independently from DT. OT should be conducted separately from DT; however, DT and OT may be combined when clearly identified significant cost/time benefits would result or separation would cause delay involving an unacceptable military risk or unacceptable increase in acquisition cost. The determination of unacceptable military risk or unacceptable increase in acquisition cost is a function of the decision review.

6. Materiel developers will design an alternative test strategy when the opportunity exists to exploit success, reduce test resources and shorten acquisition lead time. Such design should be done in coordination with the operational tester, the combat developer and the logistician. Updated development plans will include, at Milestones II and III, if appropriate, an alternative test strategy with attendant acquisition risk.

The alternative will be explained in the Coordinated Test Program or in forwarding correspondence in sufficient detail to assist the Army System Acquisition Review Council or the In-Process Review in selecting the better strategy.

7. Determinations not to conduct planned DT and OT (as outlined in the approved Decision Coordinating Paper) for major systems will be reflected in a revised Decision Coordinating Paper (DCP) and the Coordinated Test Program upon updating for approval (waiver) by DA and DOD (per DOD Directive 5000.3). Determinations not to conduct planned DT and OT for designated non-major systems, whose IPR decisions are approved by the Deputy Chief of Staff for Research, Development and Acquisition (DCSRDA) for DA, will be reflected in the In-Process Review minutes and the updated Coordinated Test Program for approval by the Under Secretary of the Army upon staff coordination. Authority to determine required testing is delegated for other non-major systems to the assigned materiel developer and operational tester for DT and OT, respectively, in coordination with the In-Process Review participants.

8. The Department of the Army System Coordinator (DASC) will, prior to the preliminary ASARC review, arrange for an ad hoc directorate level staff review within ODCSRDA, Office Deputy Chief of Staff for Logistics (DCSLOG) and Office Deputy Chief of Staff for Operations (DCSOPS) of test documentation, including the updated CTP, the independent evaluation reports for DT and OT and the test and evaluation section of the DCP. This staff review will address the adequacy of past tests, test results and evaluations, planned tests and any alternative test strategy. Review results will be provided to DCSRDA, DCSLOG and DCSOPS prior to the

time of the preliminary ASARC review. The DASC will be assisted by Office representatives of DCSRDA, DCSOPS and DCSLOG and, as necessary, representatives of OTEA, TRADOC and DARCOM.

9. The developer's independent evaluator for DT and the independent evaluator for OT and the TRADOC COEA will provide presentations to the preliminary ASARC review or the IPR, addressing test adequacy, test results, unresolved critical issues, and future testing, to include an alternative test strategy and associated risks when appropriate. Such presentations will be brief and highlight significant information in the independent evaluation reports needed for decision making.

Typical test "footstones" associated with each of the operational test cycles are given in TABLE 1. The lead/responsible organization is given in parenthesis with each footstone where it could be uniquely identified. The time periods given (based on time zero equals start of test) are not fixed but represent a general target date.

Mission Definition

It is a fundamental requirement of the methods recommended in this study project that a clear and unambiguous statement of the mission of a system be obtained. This definition should contain:

1. A description of the purpose of the system.
2. System quantitative requirements, and system critical issues.

Resources

Resources usually evidence themselves as a practical constraint on the operational test and evaluation of a system. There are four principal areas of consideration here:

TABLE 1. TYPICAL FOOTSTONES ASSOCIATED
WITH AN OPERATIONAL TEST

FOOTSTONE	TIME PERIOD
Outline Test Plan (OTEA)	1st TSARC after Milestone 0
Operational Issues and Criteria for Test (TSM)	T-270 days
Independent Evaluation Plan (TM)	T-240
Test Support Packages Furnished to OTEA	T-210
(1) Training Support Package (TSM)	
(2) Maintenance Support Package (PM)	
(3) Doctrine and Organization Support Package and Test Scenarios (TSM)	
(4) Threat Support Package (TSM)	
Complete Test Design Plan Coordination (OTEA)	T-180
(1) PM	
(2) TSM	
(3) Logistician (LEA)	
Assign Test Team (TM)	T-120
Assignments for Test	T-90
(1) Test Director (TM)	
(2) Deputy Test Director for Training (TSM)	
(3) Deputy Test Director for Doctrine (TSM)	
(4) Deputy Test Director for Logistics	
(5) Test Units	
Test Director on Site	T-60
Begin Training Cadre	T-60
Provide OT Readiness Statement (PM)	T-45
Safety Release Statement (PM)	T-45
Pretest Training (TM)	T-30
Provide OT Readiness Statement-Training, Logistics & Doctrine (TSM)	T-30
Detailed Test Plan Completed (TM)	T-30
Start Test (TM)	T-0
Complete Test (TM)	T-C
Scoring Conference (TM)	C+10
Test Report Completed (TM)	C+30
Independent Evaluation Report Completed (TM)	C+90
IPR/ASARC	C+120

1. Budget
2. Personnel resources
3. Environmental factors
4. Technological factors

These are normally addressed in the Outline Test Plan.

System Description

System description consists of either:

1. Identification of alternative system configurations
2. Configuration documentation followed by
3. A system summary description

During the conceptual phase, steps 1 and 3 form a logical sequence.

In the late Validation Phase and Full Scale Engineering Development Phase, the emphasis will increasingly shift to steps 2 and 3.

The object of the last step is to present an uncluttered picture of only those features of the system structure which have a direct bearing on:

1. The estimation of operational effectiveness
2. Military utility tradeoff study

Figures of Merit

A figure of merit is a statement which relates the program objectives to quantitative system requirements. It is a statement of the ability of a system to meet an operational need, including the recognition of the risk and uncertainty that are fundamental characteristics of the military mission.

The most comprehensive figures of merit have been dubbed operational effectiveness and military utility. Operational effectiveness is a quantitative measure of the extent to which a system may be expected to achieve a set of specific mission requirements. It is regarded to be a function of:

1. Operational suitability
2. Operational dependability
3. Operational capability

Military utility is a measure of the value received (overall effectiveness) for the resources expended (time, materiel, personnel and cost). These concepts are discussed in more detail in Appendix B.

Specification of Accountable Factors

As a preliminary to the system model construction and following mission definition system description, and specification of figures of merit, it is necessary to spell out the boundary conditions of the analysis to be conducted. First, the level of accountability must be specified:

1. What are the system interfaces?
2. What is the depth of the analysis?
3. What are the variables of the analysis?

Second, it is necessary to define constraints on:

1. Data
2. Schedule
3. Burden
4. Resources

5. Acceptable risk and uncertainty

6. Operational environment

In addition, it is necessary to spell out the accountable factors in the areas of:

1. Personnel

2. Procedures

3. Hardware

4. Logistics

Identify Data Sources

The detailed structure of the system model must be tailored to fit the type of data available. This is, of course, a two way road: only those questions may be answered for which data exists. Early identification of data sources will permit an investigation of the limitations of the expected data sources and will alert management to the necessity of planning to acquire supplementary data. This data may be derived by analysis or obtained in tests (contractor, development and/or operational tests).

Model Construction

Model construction is a four step process:

1. List assumptions

2. List variables and define model parameters

3. Construct effectiveness model(s)

4. Construct cost model(s)

The listing of assumptions is crucial. The usefulness of a model can be severely restricted if the assumptions violate reality. A clear

statement of assumptions is, therefore, a necessity in judging the validity of the results of a model exercise.

Listing variables and defining the model parameters permits a comparison of the structure of the model with the list of accountable factors. It provides a means of judging the completeness of the model structure.

Operational effectiveness models should reflect the three major system attributes:

1. Operational suitability
2. Operational dependability
3. Operational capability

Data Acquisition

Planning for data acquisition requires careful attention to:

1. Specification of data elements
2. Specification of test methodology
3. Specification of a data collection system

The key to an adequate data acquisition program is the determination of those system events which are significant. A system event is only of significance if it contributes to the evaluation of parameter of the system model. Data elements are only significant if they uniquely locate the system event in space and time with respect to other system events.

Frequently it is necessary to answer questions which call for special testing. Maximum utilization of the acquired data can be achieved only if the specification of test methodology is accomplished in a manner responsive to the needs of the development and operational testers. During

model construction any special testing that may be required should be communicated to those responsible for planning for data collection.

In the total context of this study, specification of a data collection system requires a consideration of "data" in a broader sense than its use in "data element" above. A data collection system is the organized process used to gather, store, retrieve, display, publish, and distribute a wide spectrum of system-related information including, for example, training manuals, program plans, management summaries, cost data, performance data, etc.

Data Processing

The processing of operational test data for most major Army systems is a large undertaking requiring careful attention to:

1. Parameter estimation methods
2. Administrative organization
3. Personnel selection and training
4. Software development
5. Hardware specification (computing facilities)

The specification of parameter estimation methods is a crucial step. The scope of data processing is so large that it is unreasonable to assume that those who process data are aware of all the ramifications of their work.

1. Specification of effectiveness parameter estimation methods
2. Specification of cost estimating relationships.

In recognition of the complexity of the data acquisition and data processing tasks. This study recommended the establishment of a System Information Bank for each major Army system and a System Effectiveness Information Central as a focal point for operational effectiveness information retention on an Army wide basis.

Specify Schedule

Schedule is viewed as a constraint. It is assumed that schedule control will be maintained by some form of PERT. In addition, schedule should be accounted for (possibly implicitly) in the operational effectiveness/military utility models.

Model Exercise

There are two principal uses of models:

1. Evaluation of current status
2. Prediction of potential status

Evaluation provides:

1. Surveillance of current system status against quantitative operational requirements.
2. Feedback upon the efficiency of the management decision process
3. A means of determining mission weaknesses or potential problem areas
4. A point estimate of operational effectiveness which includes all relevant factors within one conceptual framework.

Prediction provides decision aids through criticality evaluations and/or analysis:

1. Identify mission critical areas
2. Alternative system configurations
3. Problem solutions

The use of a system model involves eight steps:

1. Perform checks on model
2. Quantify figure's of merit (test data and/or calculate)
3. Do trade-offs within constraints
4. Compare results with a standard of reference
5. Calculate effect of risk
6. Calculate effect of uncertainty
7. Calculate parameter sensitivity curves
8. Interpret runs

Some elaboration on "perform checks on model" might be appropriate here as these thoughts permeate the entire study project. They consist of a set of checks on:

Assumptions: All assumptions required for the model should be explicitly stated and, if possible, supported by factual evidence. If no such evidence exists, it is advisable to state the reason for the assumption (e.g., mathematical expediency) in order to indicate the degree to which the assumptions will require further justification, and to pinpoint the areas in which errors might be introduced.

Adequacy: A model must be adequate in the sense that all major variables to which the solution is sensitive are quantitatively considered. Many of these variables will have been preselected. Through manipulation of the model, some of the variables may be excluded or restricted, and others may be introduced.

Representativeness: Although no model can completely duplicate the "real world," it is required that the model reasonably represent the true situation. For complex problems, this may be possible only for sub-parts of the problem, which must be pieced together through appropriate modeling techniques. As an example, analytic representation may be possible for various phases of a complex maintenance activity. The outputs from these analyses may then be used as inputs to a simulation procedure for modeling the complete maintenance process.

Risk and Uncertainty: The various types of unknowns involved in the problem cannot be ignored, nor can they be "assumed" out; they must be faced squarely. There may be technological uncertainties involved with some of the system alternatives, operating uncertainties involved with planning and carrying out the mission, uncertainties about enemy strategy and action, and statistical likelihoods governed by the laws of chance (referred to as risk). The simplest approach on uncertainties is to make "best guesses," but this may lead to disastrous results, since the probability of guessing correctly for every uncertainty is quite small. For cases involving statistical likelihood, the application of stochastic principles and simulation techniques may be used. For the other types of uncertainties, the general approach is to examine all major contingencies and compute resultant operational effectiveness parameters.

Validity: It must be recognized that models will not be exact replicas of the "real world." Accordingly, they should not be used blindly. Portions of every model are usually common to previously used models and can be related to quantitative knowledge of trends available from past experience. The model is validated by checks in as many familiar regions

as possible. The mode is also checked for sensitivity of its output to changes in its basic structure. These sensitivity checks are made in all areas where simplifications have been made from the "real world" case or where anomalies have resulted from the validation checks.

Certain questions will disclose weaknesses that can be corrected:

1. Consistency - are results consistent when major parameters are varied, especially to extremes?
2. Sensitivity - do input-variable changes result in output changes that are consistent with expectations?
3. Plausibility - are results plausible for special cases where prior information exists?
4. Criticality - do minor changes in assumptions result in major changes in the results?
5. Workability - does the model require inputs or computational capabilities that are not available within the bounds of current technology?
6. Suitability - is the model consistent with the objectives, i.e., will it answer the right questions?

Given that the structure of the model has been verified, the figures of merit may then be determined, and tradeoff studies made within constraints. The object of tradeoff studies is system optimization.

A model exercise is the rational basis for optimization within constraints.

Prepare Management Summary Reports

The purpose of a management summary report is to communicate/coordinate the results of a model exercise with the other test data users and to those who are responsible for making decisions, when applicable. Hence, it must be executed in a manner that increases the knowledge of the system as well as aiding the decision process. The management summary report must contain not only the main results of the model exercise in a format that is readily understood, but in addition, it should contain:

1. Data input summary
2. System quantitative requirements
3. Current system status
4. Resources (remaining)
5. Trends
6. Optimum (re)allocation of resources
7. Risk and uncertainty qualification

SECTION VI

CONCLUSIONS

The US Army does not have a systematic way of meeting the requirements cited in AR 71-3:

". . . to estimate the prospective system's military utility, operational effectiveness, and operational suitability. . ."

This report outlines a process by which these parameters may be estimated and a clear distinction is made between operational effectiveness and military utility. This process does not conflict with existing policies but supplements them.

The techniques outlined in this report are applicable to most problem areas where the systems engineering approach is appropriate and there exists a degree of subjectivity and uncertainty. In particular, certain parameters in the suitability and capability categories may be determined in Development Test. By standardizing the approach and terminology, identification of areas of mutual test interest may be accomplished and duplication avoided.

The implementation of this process should be accompanied by thorough documentation and a plan to review the results on a yearly basis. This review should consist of comparing the estimates with actual results from tests and field experience. Adjustments to the process can then be made and documented to further improve the predictions. Although there are immediate advantages to implementing this process by way of providing

a comprehensive approach to testing, the more long range benefits will start to accrue in the five to ten year span in greatly improved techniques for predicting operational performance based on limited test results.

As alluded to above, this approach provides the foundation for an evolutionary development of a consistent process to give the manager an ever improving data base to use in formulating decisions involving risk. Ways of improving the information exchange between program team members are given and suggested actions are cited. The model provides a plan for a logical development of a data base. The technique discussed in Appendix B sets forth a unified approach to evaluating the system characteristics. For those characteristics that cannot be quantitatively evaluated in tests or measurements, Appendix C cites an approach (the Delphi technique) to subjective characteristics and Appendix D gives a method of quantifying these parameters.

SECTION VII

RECOMMENDATIONS

It is recommended that this technique be established as Army policy by incorporating the concepts given in this report into an updated AR 71-3, Force Development, USER TESTING. Definitions and concepts should be incorporated/changed to agree in the following documents: AR 70-10, Research and Development, TEST AND EVALUATION DURING DEVELOPMENT AND ACQUISITION OF MATERIEL; AR 310-25, DICTIONARY OF UNITED STATED ARMY TERMS; AR 702-3, ARMY MATERIEL RELIABILITY, AVAILABILITY AND MAINTAINABILITY (RAM); and DA PAM 70-21, RESEARCH AND DEVELOPMENT, THE COORDINATED TEST PROGRAM (CTP).

It is further recommended that the US Army Operational Test and Evaluation Agency adopt this approach to operational testing.

This approach could also be utilized in the curriculum employed in the Program Manager's Course, DSMC, since it is compatible with that used in the US Air Force and US Navy. It is therefore recommended that it be considered for inclusion into the PMC.

APPENDIX A

LIST OF ABBREVIATIONS

AMSAA	US Army Materiel Systems Analysis Activity
ASARC	US Army Systems Acquisition Review Council
CDEC	Combat Developments Experimentation Command
COEA	Cost and Operational Effectiveness Analysis
CTP	Coordinated Test Program
DA	Department of the Army
DARCOM	US Army Materiel Development and Readiness Command
DASC	Department of the Army System Coordinator
DCP	Decision Coordinating Paper
DCSLOG	Deputy Chief of Staff for Logistics
DCSOPS	Deputy Chief of Staff for Operations
DCSRDA	Deputy Chief of Staff for Research, Development and Acquisition
DODD	Department of Defense Directive
DSARC	Defense Systems Acquisition Review Council
EDT	Engineer Development Test
FDTE	Force Development Testing and Experimentation
FOE	Follow-on Evaluation
FORSCOM	US Army Forces Command
FY	Fiscal Year
IOC	Initial Operational Capability

LIST OF ABBREVIATIONS (Continued)

IPR	In-Process Review
LEA	Logistics Evaluation Agency
LP	Limited Procurement
LRIP	Low Rate Initial Production
MENS	Mission Element Need Statements
ODP	Outline Development Plan
OMB	Office of Management and Budget
OT	Operational Test
OTEA	Operational Test and Evaluation Agency
PM	Program Manager
PMO	Program Manager's Office
RAM	Reliability, Availability and Maintainability
TCATA	TRADOC Combined Arms Test Activity
TEMP	Test and Evaluation Master Plan
TIWG	Test Integration Working Group
TM	Test Manager (OTEA)
TRADOC	US Army Training and Doctrine Command
TRASANA	Training and Systems Analysis Agency
TSARC	Test Schedule and Review Committee
TSM	TRADOC Systems Manager

APPENDIX B

PROPOSED TECHNIQUE FOR DETERMINING OPERATIONAL EFFECTIVENESS AND MILITARY UTILITY

The mission of the US Army Operational Test and Evaluation Agency as stated in OTEA Memo 10-1 dated 1 Oct 76, is to

"...support the materiel acquisition and force development processes by exercising responsibility for all operational testing (OT) and by managing Force Development Testing and Experimentation (FDTE), and joint user testing for the Army."

AR 71-3, dated 17 Mar 75, further states

"OT is that test and evaluation conducted to estimate the prospective systems military utility, operational effectiveness, and operational suitability (including compatibility, interoperability, reliability, availability and maintainability (RAM) and supportability, operational man (soldier), machine interface and training requirements), and need for any modifications."

However, the Army has made no distinction between operational effectiveness and military utility in reporting the results of tests.

The test and evaluation community have addressed the operational suitability and the need for many modifications in the test reports and evaluations. There still exists a need for a technique to address the military utility and operational effectiveness of the system. There is also a need for an effective process for identifying the critical areas for test and evaluation that will permit the independent evaluator to answer the proper questions. It should be noted that under ideal conditions, the selection of critical issues is properly a US Army Training

and Doctrine Command function but with an admitted overview and co-ordinated concern on the part of the operational tester. However, the technique described in this note may be employed by the developer, tester and user. The results, after becoming familiar with the approach, will probably be a significantly improved product and at the same time make better use of our limited scientific and technical resources. It will also help highlight the true critical areas for test design and field test. It could further serve as a catalyst to open the door with TRADOC/TRASANA to a methodology for greater benefit to decision makers as well as evaluators.

At this point it is appropriate to define some of the terms with which this Appendix will be dealing.

SYSTEM is an arbitrary collection of physical configurations together with the functions performed upon them. It is completely defined, when and only when, a set of influencing configurations and functions, called the background, is given.

PERFORMANCE is the sub-set of all system outputs which relate to the requirements.

SYSTEM EFFECTIVENESS is a measure of system performance that is evaluated over a specific background where the measure is representative of the degree of correspondence between performance and requirements.

MILITARY UTILITY (from AR 310-25) "is the military/operational value of an item/system when measured from within a pertinent Army Concept Program and against the threat analysis and future concept, doctrine, environment, organization, skills, availability, reliability, maintainability, obsolescence and other materiel objectives/requirements."

If the background is the operational field environment, then system effectiveness becomes operational effectiveness. The operational effectiveness, therefore, is a measure of the ability of a system to accomplish a particular function in an operational environment. Military utility is the military worth of a system performing its mission in a tactical (competitive) environment including the versatility (or potential) of the system. It is seen that the operational effectiveness of a system is a major influencing factor in the utility of a system. However, when comparing two systems, say anti-tank weapons, where both are of equal operational effectiveness, and one is effective against troops where the other is not, then the military utility of the one including the troops capability is higher.

Therefore I propose to redefine operational effectiveness and military utility as follows:

OPERATIONAL EFFECTIVENESS is how well the system performs in a particular (intended) mission in an operationally competitive environment. Operational effectiveness estimates form a basis for judging the adequacy of our defense posture.

MILITARY UTILITY is the system worth when applied to a particular function. Military utility estimates form a rational basis for making management decisions.

Safety is a paramount consideration for many weapon systems. Unless safety features are carefully considered during the development phase, there is a significant probability that a system may be activated by error (operator, maintenance, spurious signals, failure of a critical circuit or function, etc.). Military and strategic consequences of such errors are enormous, and their prevention is frequently an overriding factor in the evaluation of military utility.

Security is a vital factor for some systems. What is the probability that a saboteur could take over a system and render it incapable of use (Norway - WW II), or worse, use it against us? Although it may be difficult to quantify both, safety and security, there can be no question that system design and operational criteria must reflect a thorough assessment of these real probabilities, and due consideration should be given to their inclusion in developing critical issues and determining the scenario.

Also important to the technique being developed are the following definitions:

OPERATIONAL CAPABILITY is the measure of the results of the mission; given the condition of the system during the mission (dependency).

OPERATIONAL DEPENDABILITY is the measure of the system condition during the performance of the mission; given its condition (suitability) at the start of the mission.

The relationship of operational effectiveness can best be shown diagrammatically as in Figure 1B and that of military utility in Figure 2B (note that the lists are not exhaustive).

Adopting this technique to analyze the system allows the user, developer and operational tester to better assess the areas of influence overlap.

Actions - PM, TSM and TM should consider this approach to analyzing the system to determine those areas that may provide data to satisfy the needs of both the developer and operational tester. Also it may be remembered that with all the scientific and engineering improvements in weapons and organization, one must not lose sight of the unchanging fact that no matter how fine the weapons and equipment of the Army, the effectiveness with which they are employed depends ultimately on the skill, the intelligence, the courage, and the dedication of the men who use them. Therefore, military judgment will be a major factor in quantifying both operational effectiveness and military utility.

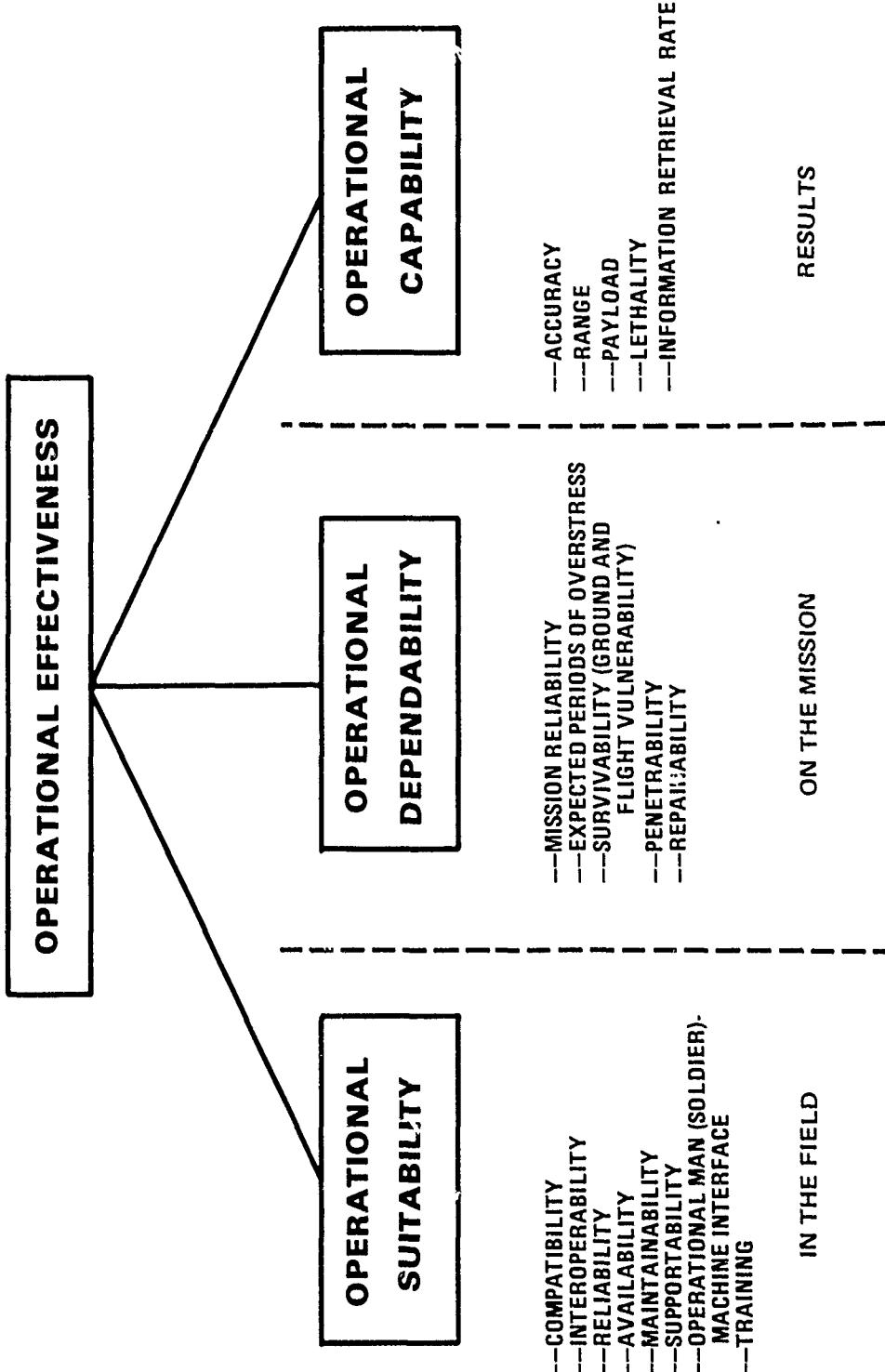


Figure 1B. Analytical Approach to Operational Effectiveness.

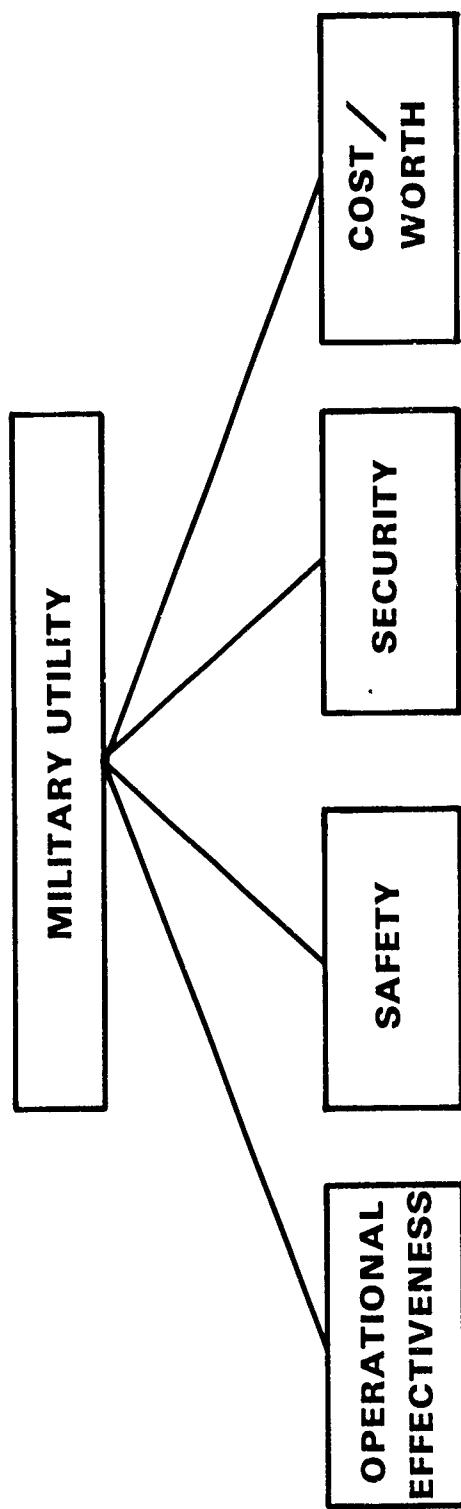


Figure 2B. Synthesis of Military Utility.

APPENDIX C

DELPHI TECHNIQUE

Background

The "Delphi Technique" is a term referring to the procedures for obtaining and refining the opinions of a group of people. The technique was developed by the oracles of the city of Delphi on the slopes of Mount Parnassus in Phocis, Greece in ancient times. The technique was further refined by the RAND Corporation between the years 1948 and 1959. It is currently being used by government and industry for long range planning.

Problems With Opinion

A basic characteristic of opinions, as opposed to more solid knowledge, is the fact that if you interrogate several equally competent individuals, you are likely to get a divergence of answers. From the point-of-view of the decision maker, a divergence of estimates creates a problem of how to use the estimates in fashioning his policies. Traditionally, decision makers have either selected single advisors, a staff, or appointed committees to provide them with expert opinion. There is always an obvious danger associated with limiting oneself to the opinion of just one advisor. Use of groups also has its shortcomings.

One problem with group opinion is the influence of the dominant individual. A very convincing series of studies have shown that the group opinion is likely to be highly influenced, if not determined, by the views of the member of the group that does the most talking or has the highest military rank; there is no significant correlation between success in influencing the group and competence in the problem being discussed. Another difficulty is the frequent introduction of irrelevant and/or redundant materiel that obscures the directly relevant materiel offered by participants. A third difficulty is the group pressure that over values "reaching a consensus."

Procedure

The Delphi procedure has been designed to reduce the effects of the above undesirable aspects of group interaction. The procedure has three distinctive characteristics: anonymity, controlled feedback, and statistical group response.

Anonymity is a device used to reduce the effect of the dominant individual. It is maintained by eliciting separate and private answers by written questionnaires. All interactions between respondents is through formal communication channels controlled by the administrator.

Controlled Feedback is a device to reduce the introduction of irrelevant and/or redundant materiel. A Delphi exercise will usually consist of several iterations where the results of the previous iteration are fed back to the respondents, normally in summarized form.

Statistical Group Response is some form of statistical index representing the groups opinion. For cases where the group task is to estimate a numerical quantity, the median of individual estimates has been found to be the most useful index. Consequently, there is no particular attempt to arrive at unanimity among the respondents, and a spread of opinion on the final round is the normal outcome.

APPENDIX D

UTILITY THEORY

Concept

Utility is a personal subjective valuation of a tangible or intangible.

Definition

Utility Theory is a set of loosely defined lemmas associated with peoples preferences, values, and/or assumptions about a person's preferences, that permits their systematic valuation in numerically useful ways.

Essentials

A basic lemma is: if a decision maker is indifferent between two alternatives, the expected utility of the alternatives is the same. Another is: people make decisions to maximize expected utility rather than monetary value. The following are essentials of the theory:

- a. A set X of elements x, y, and z, ...called decision alternatives or courses of action.
- b. An individual's (or group's) preference-indifference relationship.
- c. A set of internally consistent assumptions about X and how the preference-indifference relationship behaves on X.

d. Any theorems which can be deduced from the assumptions. For additional information on decision making under uncertainty, refer to a good statistic's book, e.g., References 6 or 7.

Churchman-Ackoff Theory

This theory (developed in Reference 8) is based on ordering potential outcomes and then assigning numbers to the outcomes which reflect their relative values. That is, outcomes are ranked on an ordinal scale which is then converted to an interval scale by assigning values to the outcomes. This will be accomplished by the Delphi Technique as outlined above. This theory implies additivity of relative values and assumes a linear utility function. For illustration, consider a situation which has five available alternatives (or outcomes). In applying the Churchman-Ackoff method, the following steps are performed:

- a. Rank the outcomes in decreasing order of importance, $F(1)$, $F(2)$, $F(3)$, $F(4)$... $F(n)$.
- b. Determine which is preferred, $F(1)$ or the combination of $F(2)$, $F(3)$, $F(4)$... $F(n)$. If the combination is preferred, then determine which is preferred $F(1)$ or the combination $F(2)$, $F(3)$, $F(4)$... $F(n-1)$. If the combination is still preferred, continue to drop outcomes until $F(1)$ is preferred to the combination.
- c. Determine which is preferred, $F(2)$ or the combination $F(3)$, $F(4)$... $F(n)$. If the combination is preferred, then drop outcomes until $F(2)$ is preferred (as in step b).

d. Determine which is preferred, $F(3)$ or the combination of $F(4) \dots F(n)$. If the combination is preferred, then drop $F(5)$ and so on until completion.

e. Assign numbers (or ratios) to the outcomes which reflect their relative rankings as determined above.

EXAMPLE: In the evaluation of the SAM-X weapon system, the following factors are deemed important and need some type of quantification for an analysis: (1) Kill probability given a mission, (2) Cost to prevent threat/cost to maintain threat , (3) Fraction of total targets effective against (aircraft, missiles, etc.), (4) target worth/munition worth (0.5), and (5) Probability of system surviving.

Assume the factors are ranked in their order of importance and call them $F(1)$, $F(2)$, $F(3)$, $F(4)$ and $F(5)$. Now see if:

$$F(1) > F(2) + F(3) + F(4) + F(5)$$

$$F(2) > F(3) + F(4) + F(5)$$

$$F(3) > F(4) + F(5)$$

If the above preferences hold, then numerical values $V(1)$, $V(2)$, ... $V(n)$ are assigned to the outcomes which are consistent with the preferences. $V(1) = 16$, $V(2) = 8$, $V(3) = 4$, $V(4) = 2$, and $V(5) = 1$.

so $V(1) > V(2) + V(3) + V(4) + V(5)$ or $16 > 8 + 4 + 2 + 1$

$$V(2) > V(3) + V(4) + V(5) \text{ or } 8 > 4 + 2 + 1$$

$$V(3) > V(4) + V(5) \text{ or } 4 > 2 + 1$$

$$V(4) > V(5) \text{ or } 2 > 1$$

The SAM-X kill probability, given a mission (it is assumed here that mission designation is independent of status of system) depends on a number of sequential probabilities--the final one being the probability of success when the missile gets to the target; the preceding one is the probability that when the button is pushed, the system performs its intended program in getting the missile to the target area; the initial one is the system operational suitability or the probability that the system is operating and/or ready to function when the emergency is at hand. The SAM-X is an air defense system against air breathing aircraft. It has practically no anti-missile capability. It is large, costly, technology of fifteen years ago and designed to meet a threat that may have substantially changed. The SAM-X has a good probability (0.8) of killing an air breathing aircraft in spite of holes in the tracking pattern (blind spots). The cost to prevent the threat is minimal (5 million) in that existing systems will be used to detect the launching, and training in evasive maneuvering is a standard course for all pilots. Also the enemy current strategy calls for the use of missiles in those areas where the SAM-X is likely to be deployed. The system is planned to be employed at thirty sites, requiring fifty men per site at a total estimated cost of five billion dollars. It is currently estimated that the percentage of targets where the SAM-X has a capability as compared to the total enemy density of air targets is thirty percent. The average target worth is about the same as the

missile worth. The system consists of five large (but mobile) subsystems of which only the launcher must be in an exposed position. However, the limitations on the separation distances (cable lengths) generally forces at least two of the five to be exposed. Camouflage equipment is provided but is not effective against radar or infra-red. The SAM-X has a readily recognizable signature and is expected to be a prime target for air-to-ground missiles. Since it is very vulnerable to attack, the Delphi group set the survival ratio as "0.6."

The military utility is now calculated as:

$$\begin{aligned} \text{Utils} &= 16(0.8) + 8(5/5000) + 4(0.3) + 2(1)(0.5) + 0.6 \\ &= 12.8 + 0.008 + 1.2 + 1 + 0.6 = 15.6 \end{aligned}$$

$$\begin{aligned} \text{Max Utils} &= 16(1) + 8(1 \text{ break even}) + 4(1) + 2(1 \text{ break even}) + 1 \\ &= 16 + 8 + 4 + 2 + 1 = 31 \end{aligned}$$

$$\text{Military Utility} = 15.6/31.0 = 0.50 \text{ Utils}$$

At the initial session, the Delphi group had agreed that a minimum acceptable ratio would be that given in Figure 1D. It is seen that a minimum acceptable ratio corresponding to five major characteristics is 0.8. Thus, it is seen that the military utility is rather low. The best place to spend money improving is increasing the kill probability and expand survival capability. The most critical test areas are the kill probability and survival capability. Since the threat ratio is critical to the equation (weight of 8) and is so small, it appears that the system should be sent back to R & D for basic changes in direction

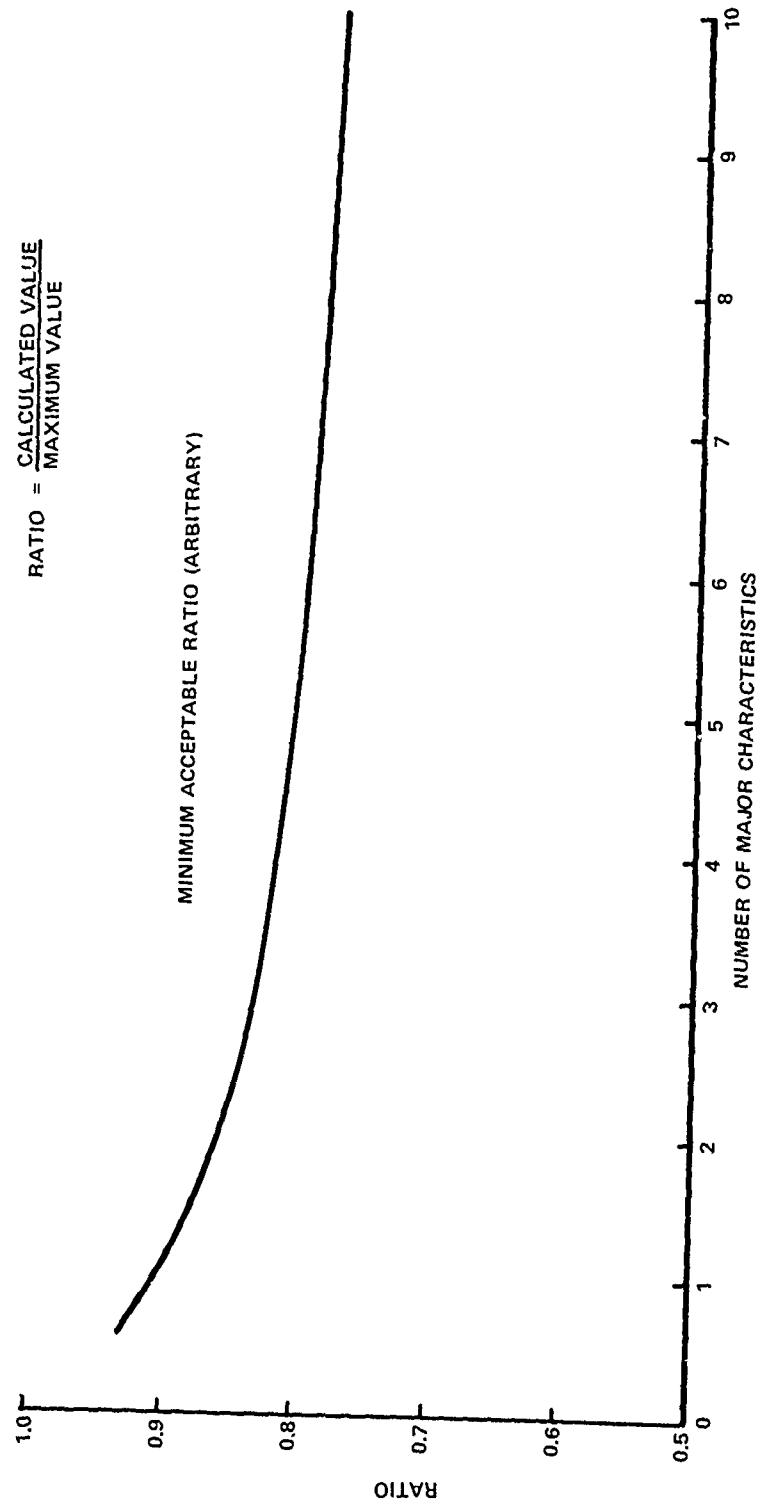


Figure 1D. Minimum Acceptable Ratio for Military Utility.

(the high cost of kill, can kill you!).

Applications To Test and Evaluation

The possible applications of utility theory to test and evaluation functions appear to be many. Particularly in trying to access/evaluate the military utility of a system or in comparing various systems. No attempt is made here to develop utility in its many highly technical aspects but merely to show a possible technique to employ in the weapon system life cycle. It is assumed that the professional engineer has been exposed to utility theory and has access to literature such as Reference b. The details, characteristics and preferences are expected to vary widely but the effectiveness in thought orientation, ability to quantify judgment, and the combining of the knowledge of many will prove invaluable during the early life cycle of the system.

It is envisioned that after a Mission Element Need Statement (MENS) has been established and prior to the operational/critical test issues being established, an OTEA test manager will be appointed. This system manager would study the characteristics of the system and select from six to ten experts in the various areas involved in the system function. The selection will be reviewed by the Scientific Advisor (for technical expertise) and the Commanding General (for military expertise). On approval this group would constitute the Delphi forum. The initial task would be to establish the operational/critical test issues of the new system and to order them in accordance with priority (top to bottom).

Each major characteristic would be structured into subgroups (e.g., kill probability - probability of a destruct or mission abortion). These subgroups would then be developed, structured and quantified by combining the Delphi technique and utility theory. This type of sensitivity analysis would be used as the input to the design of the operational test. It will be necessary for the forum to open with a detail briefing by the intelligence section on the current status of the threat (this will become an official part of the record). As the contractor and development test results become available, the numerical values would be refined. If it becomes apparent that a reorientation is needed, the forum will be reconvened.

At the end of each Operational Test, the forum will be given an updated estimate, and this would be used to re-evaluate the positions by the Delphi Technique. The evaluation of military utility will also be direct fallout of this process.

Conclusions

It is seen that by combining the Delphi Technique and utility theory, a formal method may be developed to better quantify the best judgments available to OTEA and at the same time give each member an exceptionally good insight to the total program. The results will be valuable to the test designer and the evaluator. As the process is employed, it is expected that many innovations will greatly increase the value and extend the applications.

Recommendations

It is recommended that a system in the early part of the life cycle be selected for the application of this technique on an experimental basis. A report on the results, modifications, recommendations, and attitudes toward the technique should be accomplished after each milestone.

LIST OF REFERENCES

Reference No.

- 1 "Review of Army Analysis and Test and Evaluation Agencies," report dated 11 May 1977, Office Chief of Staff, Army.
- 2 "Weapon System Effectiveness Industry, Advisory Committee-Chairman's Final Report," report No. AFSC-TR-65-6, dated January 1965, USAF, Air Force Systems Command.
- 3 "Analyzing Training Effectiveness," report No. TRADOC PAM 71-8, dated 1 December 1975, USA, Training and Doctrine Command.
- 4 "Total System Management," DRAFT report dated 3 May 1977, USA Training and Doctrine Command.
- 5 "Operational Test and Evaluation Methodology Guide," report dated May 1976, USA Operational Test and Evaluation Agency.
- 6 "Statistical Analysis for Business Decisions," book by W. A. Spurr and C. P. Bonini, Revised Edition 1973, published by Richard D. Irwin, Inc.
- 7 "Management Decision Making Under Uncertainty," book by Dyckman, Smidt and McAdams, published by Macmillan Co., London, England.
- 8 "Scientific Method, Optimizing Applied Research Discussions," book by R. L. Ackoff, John Wiley & Sons, 1962.